

Effective image data points extraction via decision tree based on feature classifier

C. S. Sasireka & R. Santhosh

To cite this article: C. S. Sasireka & R. Santhosh (2019) Effective image data points extraction via decision tree based on feature classifier, *Automatika*, 60:5, 519-525, DOI: 10.1080/00051144.2019.1576966

To link to this article: <https://doi.org/10.1080/00051144.2019.1576966>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Apr 2019.



Submit your article to this journal



Article views: 1075

[View related articles](#) View Crossmark data 



Effective image data points extraction via decision tree based on feature classifier

C. S. Sasireka and R. Santhosh

Department of Computer Science and Engineering, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore, India

ABSTRACT

Image data points mining is concerned with the extraction of knowledge relationship among image data and other patterns inherent in the images. Taxonomy-Aware Catalog Integration (TCI) processing step ensures that the master taxonomy Rule-based Multivariate Text Feature Selection (RMTFS) method takes into consideration both the semantic information and the syntactic relationships between n-gram features. In realizing the relationship between the attributes, the Decision Tree-based Label Feature Classifier (DTLFC) mechanism is proposed. In the initial step in the DTLFC mechanism, pixel-wise image feature points are extracted and the same is converted into a table of a database. The feature descriptor thus formed by combining a features set and a specific pixel's label is represented by a tuple of the transformed database table and helps in the generation of the decision tree using a given image points data set in constructing a pixel-wise image processing model. Both experimental and theoretical analysis demonstrate that the DTLFC mechanism can attain a very efficient and effective level of classification on image data points. Compared with the TCI and RMTFS methods, the DTLFC mechanism creates an effective feature classifier in terms of filtering efficiency, classification accuracy, processing time, computational cost, scalability, and the intensity rate of dominant attributes.

ARTICLE HISTORY

Received 23 November 2018
Accepted 23 January 2019

KEYWORDS

Text feature selection; decision tree; classification accuracy; image data points; feature classifier; domain attributes; label feature set

1. Introduction

Image mining is concerned with inherent information extraction, image data association, or other patterns possibly ambiguously embedded in the images and draws leading capability from computer vision, image understanding, data mining, machine learning, database, and artificial intelligence. Large databases with multiple heterogeneous data sources require new ways to treat them. In this case, data mining techniques implemented independently or coupled with different other techniques attempt to discover the implicit knowledge present in a data set. These techniques essentially aim to search image data points, to depict contents as well as extract information in a meaningful way. The information that exists in organizations is familiar and shapeless, and these techniques also address the textual and multimedia beyond digital data and evidence.

In this work, the Decision Tree-based Label Feature Classifier (DTLFC) mechanism initially extracts the pixel-wise image feature points. The feature points are placed in a database table and each tuple in the table has a feature descriptor. The feature descriptor contains the label of a particular pixel with a feature set. The feature descriptor in the DTLFC mechanism denotes a piece of information that is relevant to solve a certain class

of application problems. Based on diverse local pixel information, different image processing techniques are applied at different image locations within a single scan image of the test Single Proton Emission Computed Tomography (SPECT) Heart data set.

2. Literature review

Textual and multimedia data sampling using an online oversampling principal component analysis (osPCA) algorithm in [1] detects the existence of outliers from a huge quantity of data via online updating. The osPCA algorithm with a multi-clustering structure fails extremely in maintaining high-dimensional space. In handling high-dimensional data, the covariance matrix needs to be subtracted, and the principal components analysis (PCA) might not be preferable for estimating the principal directions for such kind of data. The Bayesian framework developed in [2] estimates the principal direction densities for the receiver operating characteristic curve which fails to incorporate variable complexity, which is possible through regularization and the use of the Occam factor. Popular subspace learning algorithms such as the PCA for cross-domain face modelling and the Fisher's linear discriminate analysis in [3] perform cross-domain face recognition and

text categorization. The effectiveness of the new regularization for cross-domain classification and reconstruction tasks realizes numerous transport subspace learning algorithms by introducing regularization.

Subsequent to the subspace learning process, the FAST clustering-based feature selection algorithm, as described in [4], separates features into groups of clusters using graph-theoretic clustering algorithms. The following step, the most delicate feature that is powerfully correlated to objective classes from each cluster, is selected and a subset of features is formed. The fuzzy similarity-based self-constructing algorithm for feature clustering discussed in [5], based on the similarity test, is grouped into clusters. The extracted feature, corresponding to a cluster, is a subjective grouping of the words restricted in the cluster. The closely matched derived membership functions explain the distribution of the training data. A new cluster-adaptive distance bound based on separating hyperplane boundaries of Voronoi clusters, as described in [6], complements the cluster-based index. This cluster-based index bound enables efficient spatial filtering with a comparatively minimum preprocessing storage overhead and is applicable to the Euclidean and Mahalanobis similarity measures, but it fails to retrieve and process petabytes of data for a variety of filtering purposes such as data mining and decision support.

The flexible rule-based system, as demonstrated in [7], allows users to customize the filtering criterion to be applied on walls, and a Machine Learning-based soft classifier automatically labels messages in content-based filtering. The system allowing online social network users has a direct control on the messages posted on their walls. But it fails to integrate contextual information related to the name of all the groups in which the user participates, appropriately weighted by the participation level. An adaptation of the multistep Nearest Neighbour (NN) algorithm illustrated in [8] incurs an unaffordable network overhead which requires a proof of correctness of the obtained result. The server returns the NN set along with supplementary information that permits consequence verification using the data set signature.

The Extended XML tree pattern expressed in [9] includes P-C, A-D relationships, exclusion functions, wildcards, and position restriction. The XML tree pattern investigates an academic framework about similar cross which demonstrates the essential reason in the proof of optimality on holistic algorithms. Based on theorems, a set of novel algorithms professionally processes three categories of extended XML tree patterns. Extended Subtree, as illustrated in [10], generalizes base distances by providing new rules for subtree mapping. However, the new approach does not seek to address the problems involved in lessening computational complexity and thereby improving quality.

The distributed method for detecting distance-based outliers in very large data sets, as illustrated in [11], is employed to forecast novel outliers. The temporal cost is expected to be at least three orders of magnitude quicker than the conventional nested-loop-like approach to distinguish outliers. The distance-based outlier is also useful for the parallelized version of other kinds of algorithms based on support vector machines. The ontology-based fuzzy video semantic content model, as illustrated in [12], uses spatial and temporal relatives in occasion and notion definitions. The meta-ontology description provides a wide-domain specific rule that builds normal form that allows the user to construct ontology for a given domain.

The taxonomy-aware processing step, as demonstrated in [13], which regulates the results of a spatial computation cost and a text-based classifier, makes the products that are close together in the provider taxonomy effective. The taxonomy-aware processing step prepares intuition as a prearranged prediction optimization problem but does not explore semi-supervised learning techniques to incrementally retrain the base classifier with elements chosen during the taxonomy-aware calibration step.

In [14] semi-supervised classification, an objective function is created by combining together the global loss of the local spline regressions and the squared errors of the class labels of the labelled data. In this way, a transductive classification algorithm is developed in which a globally most favourable classification is attained. In the semi-supervised learning setting, the algorithm is considered and it deals with the Laplacian regularization framework, but it does not carry out effective image segmentation and image matting.

Multiple-Instance Learning via Disambiguation (MILD), as described in [15], identifies the true positive instances for image segmentation. The underlying principle of the MILD fails to bridge the gap between multiple-instance learning and SIL.

A feature relation network (FRN) considers semantic multiple-instance and single-instance information, as shown in [16], which also leverages the syntactic relations among n-gram features. The FRN is intended to professionally facilitate the enclosure of comprehensive sets of heterogeneous n-gram features for enhanced sentiment classification. Based on the results, the FRN is suitable for other text classification problems, where semantic information is available. The FRN does not discover additional potential feature relations and does not add additional feature occurrence measurements. After an analysis of the trajectory data, along with the locations traversed by vehicles, an ordering of these locations is critical for improving the accuracy of classification. Classifying trajectories on road networks

explained in [17,18] is not efficient and effective methods for pattern-based classification. The security of pattern classifiers fails to develop techniques for simulating attacks for different applications.

After performing the pattern-based classification, a new partitioning method called “Clustering, De-clustering, and Selection” (CDS), is depicted in [19] to approximate the statistical characteristics of the training partitions. The effectiveness of the different types of training partitions, a huge number of disjoint training separations with individual distributions were generated, but this process was not effective in developing training data subsets with overlaps. The filter-based data partitioning approach with a wrapper-based method is not integrated, in which clusters with different distributions and diversities are generated.

3. Decision tree-based label feature classification

In the DTLFC mechanism, the tree size is independent of the data base size. The tuples in the database table pass through the decision tree. The DTLFC mechanism performs the classification operation across the fixed height of the tree, and decision trees are constructed for multi-attributed data on DTLFC mechanism image points. Based on different local pixel information, different image processing techniques are applied at different image locations within a single scan image of the test SPECT Heart data set. The architecture of the DTLFC mechanism is shown in Figure 1.

In the DTLFC mechanism, the SPECT Heart data set is taken as input and formatted as a set of identically sized unprocessed label image pairs. The mechanism selects important features and stores extracted data in a data base table. Image data points are processed through clear decision rules and it is applied to perform pixel-wise image processing. With decision-tree generation, the DTLFC mechanism is used to mine hidden

relationships between the target labels from the image pixels and attributes of image pixels.

3.1. Decision tree-based feature extraction

The unprocessed image is an n -dimensional light-intensity function, denoted by $P(b_1, b_2, b_3, b_n)$, where the amplitude of P at spatial coordinates (b_1, b_2, b_3, b_n) gives the intensity of the unprocessed image at that pixel. The label image of the DTLFC mechanism is an n -dimensional light-intensity function, denoted by $L(b_1, b_2, b_3, b_n)$, where the value of L at spatial coordinates (b_1, b_2, b_3, b_n) gives the class identifier of the pixel at the same spatial coordinates of its corresponding unprocessed image. The database-like table $Y = (y_1, y_2, y_3, y_n)$ is a set of records, where each record $Y_n \in P^k$ is a vector, with elements $(att1_1, att2_2, att3_3, att4_4)$ being the value of attributes of Y .

The pixel value of the unprocessed image represents the grey level of a pixel. The pixel value of the label image represents the class label of the pixel. In the DTLFC mechanism, both the pixel values pertain to the same position. A pixel of the exterior contour of the unprocessed image has the value “1” in the label image for the corresponding pixel. It is a pixel of interest in the case of the DTLFC mechanism. In reality, the pixel value of the label image might take any kind of form, including binary.

In addition, many unprocessed images and a label image pair are taken as the input in the DTLFC mechanism. By transforming a set of unprocessed and label images into a database table and allowing feature extraction algorithms to work on the table, the DTLFC mechanism can mine valuable information from them.

The feature extraction algorithm is described below

//Feature Extraction

Procedure SPECT HeartImage (image: unprocessed, label)

Begin

Step 1: Set generated feature functions $[1 \dots n]$

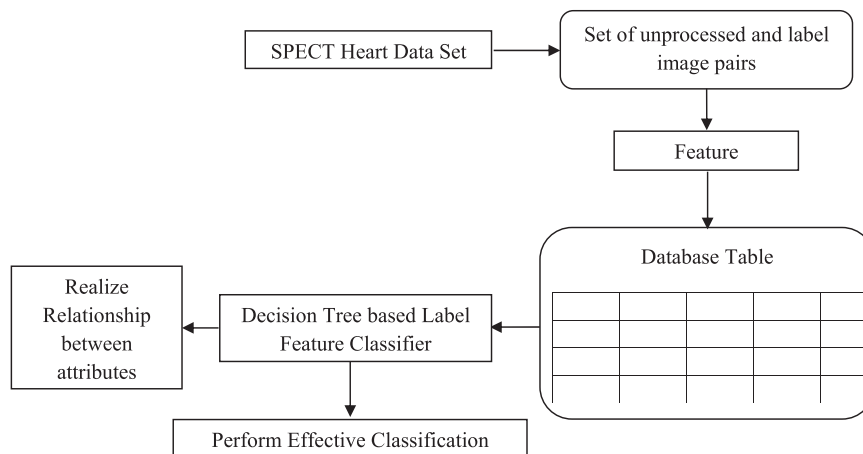


Figure 1. Architecture diagram of the DTLFC mechanism.

Step 2: Set generated label function
 Step 3: Initiate the database table with pixel points (x, y)
 Step 4: While the pixel exists, perform the pixel scanning process
 Step 5: Insert into the table value: = feature_generated1 (unprocessed, pixel)
 Step 6: Generated feature (unprocessed, pixel)
 Step 7: Generated label (label, pixel)
 Step 8: Extend the scanning process on the next pixel
 End while
 Step 9: Return to the database table
 End

As mentioned above in the algorithm, a set of features with labels are extracted. The extracted features are then placed in the database table. This table contains the pixel points (x, y) in the row and the features extracted indicate places in the column order. After the pixel scanning process, the table value is inserted. The results of the extraction process in the DTLFC mechanism help us in getting a better understanding of the image properties and enable us to relate them to real-world situations.

Other pixel-wise features like contrast, mean, and entropy are also encoded in the table on the SPECT Heart data set collected. Encoding strategies in the DTLFC mechanism such as normalization adjust the value ranging from 0 to 1 and are applied to generate the desired features. Additionally, a column in that table represents the label image point. The unknown attribute relationships among the two kinds of images (unprocessed and label) are mined using the label feature present.

3.2. Decision tree-based feature classifier

The decision tree-based label classifier is used to classify based on decision support application. In the DTLFC mechanism, decisions have to be made by the surgeon like whether the maximum item set is available in the decision tree. In the DTLFC mechanism, the decision tree classifier with association rule classification provides a better option for the surgeon to classify the benign and malevolent images. It is done by comparing the maximum items generated by the association rules in the SPECT Heart images.

As shown in Figure 2, the decision-based label feature classifier classifies the rules generated by the decision rules into normal, benign, and malevolent attributes. From a set of images, the DTLFC mechanism builds decision trees, using the concept of information entropy. The heart images from the unique client identifier (UCI) repository are a set of model "M" where

$$M = m_1, m_2 \dots, m_n \quad (1)$$

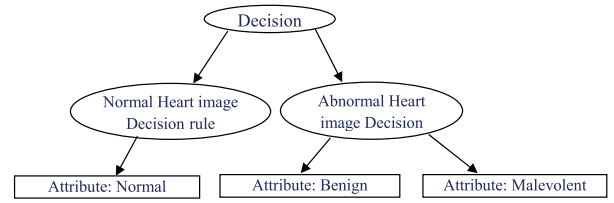


Figure 2. Decision tree-based classification of attributes.

$m_1, m_2 \dots, m_n$ are classified models with $m_i = att1_1, att2_2, att3_3, att4_4$, where $att1_1$ is the attribute of the model. The first step is to find the pixels in the x and y directions by smoothing the heart image data points from the UCI repository and take the derivative. If a two-dimensional form D_{xy} is convolved with the image S for smoothing, the operation is given by Equations (2) and (3).

$$h_x = \frac{\partial(s * D_{xy})}{\partial x} \quad (2)$$

$$h_y = \frac{\partial(s * D_{xy})}{\partial y} \quad (3)$$

The $*$ operator denotes convolution. The above form is computationally not the most efficient method to gradient computation. Using the fact that the Gaussian is separable, Equation (4) depicts a more efficient form.

$$\begin{aligned} d_x &= \frac{\partial(s * D_{xy})}{\partial x} = s * \frac{\partial(D_{xy})}{\partial x} \\ &= s * \frac{\partial(D_x * D_y)}{\partial x} = s * \frac{\partial(D_x)D_y}{\partial x} \end{aligned} \quad (4)$$

Likewise, the gradient in the y direction of the pixel in the DTLFC mechanism is calculated in Equation (5)

$$\begin{aligned} d_y &= \frac{\partial(s * D_{xy})}{\partial y} = s * \frac{\partial(D_{xy})}{\partial y} \\ &= s * \frac{\partial(D_x * D_y)}{\partial y} = s * \frac{\partial(D_y)D_x}{\partial y} \end{aligned} \quad (5)$$

The d_x and d_y give the gradient estimation in the respective (x and y) directions. In image data points, let $d_x(x, y)$ be the gradient at (x, y) in the x direction and $d_y(x, y)$ be the gradient at (x, y) in the y direction of the pixel. The magnitude of the gradient $d(x, y)$ is given in Equation (7).

$$d(x, y) = \sqrt{d_x(x, y)^2 + d_y(x, y)^2} \quad (6)$$

The processing time for the extraction of information from the database is approximated using Equation (8).

$$d(x, y) = |d_x(x, y)| + |d_y(x, y)| \quad (7)$$

The given input image data point is extracted into the number of objects from the database table. For every image from the SPECT Heart data set, the texture features have been considered and stored in the database table for further classification using the decision tree.

3.3. Algorithm for DTLFC mechanism

One rule of the data is chosen by the DTLFC at each node of the decision tree, and it most effectively splits its set of models into subsets enriched into classes. The DTLFC algorithm recurses on the smaller sub lists and the tree generation is explained through the following algorithm.

// DTLFC mechanism

Input: SPECT Heart data set, set of target keywords, set of non-target keywords

Output: Accurate classification based on decision tree

Step 1: Extract features from image data points
// SPECT Heart data set consists of records with unprocessed and label image pairs

Step 2: Target attribute achieved with the same value

Step 3: If the set of non-target keywords is empty,
then

return a root node with the most frequent value of the target attribute that is found in the set

Step 4: Attribute with the largest gain achieved among the attributes of non-target keywords

Step 5: Subsets of the main root node consist of the leaf node respectively in records

Step 6: Obtain root image points and edges and label correspondingly based on decision tree rules

Step 7: Recursively apply the decision tree classifier to the subsets until the leaf node is reached

4. Experimental evaluation of DTLFC mechanism

The DTLFC mechanism is implemented in MATLAB. The mechanism uses the SPECT Heart data set from the UCI repository for experimental evaluation. The SPECT Heart data set contains data on cardiac SPECT images. Normal and abnormal are the two categories in which patients are classified into. The associated task in the SPECT Heart data set is based on classification.

The database of 267 SPECT image sets (patients) was a progression to extract features that summarize the original SPECT images. As a result, 44 permanent feature patterns were generated for each patient. An additional processing on the pattern was to get hold of 22 binary feature patterns. The rules from these patterns were constructed using the CLIP3 algorithm, which is 84.0% accurate when compared with the diagnoses made by cardiologists. SPECT is a good data set for testing ML algorithms as well, which has 267 instances that are described by 23 binary attributes.

The DTLFC mechanism generates the decision tree to realize the relationship between attributes and the labels from image pixels. The DTLFC mechanism is compared against the Taxonomy-Aware Catalog Integration (TCI) processing step and the Rule-based Multivariate Text Feature Selection (RMTFS) method. An

experimental evaluation is made on factors such as filtering efficiency, classification accuracy, processing time, computational cost, scalability, and intensity rate of dominant attributes.

SPECT image sets from the UCI repository analyse the DTLFC mechanism through the filtering efficiency factor. The label feature classifier performs an effective filtering of each tuple in the database table, measured in terms of percentage. Classification accuracy of the different approaches report the best results over all possible choices of similarity metric in the DTLFC mechanism using decision tree. Total processing time is defined as the aggregate of time taken to extract the features, classify the features using decision tree, and realize the relationship between the attributes.

The computational cost factor of the DTLFC mechanism is defined as the equivalent state obtained through the decision tree classifier, measured in terms of milliseconds (ms). The DTLFC mechanism helps in eliminating base classifier mistakes and classifies images from dissimilar categories. The scalability of the DTLFC mechanism is the ability of a system to handle a growing amount of images to perform the classification process in a capable way to accommodate the growth in data mining, measured in terms of percentage. The intensity rate of the dominant attribute is depends on the measurement of power rate for performing the decision tree classifier.

5. Result analysis

The DTLFC mechanism is compared against the TCI processing step and the RMTFS method. The attributes considered for comparison with existing systems are filtering efficiency, classification accuracy, processing time, computational cost, scalability, and intensity rate of the dominant attributes.

Figure 3 illustrates the filtering efficiency on each training test image. In the DTLFC mechanism, the decision tree with association rule classification provides a better option for the surgeon to filter the images.

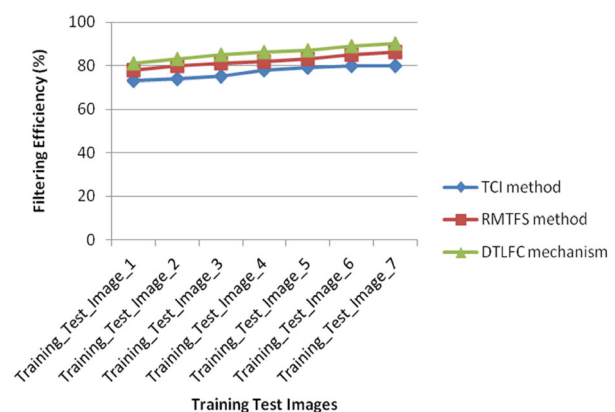


Figure 3. Measure of filtering efficiency.

It is done by comparing the maximum items generated by the association rules on the SPECT Heart images. SPECT Heart training set images produce the 10–13% improved filtering efficiency in the DTLFC mechanism when compared with the TCI method. The DTLFC mechanism is 3–5% improved filtering percentage when compared with the RMTFS method.

The classification accuracy measure is shown in Figure 4 where the accuracy rate is expressed in percentage. The classification accuracy factor has seen an improvement in the DTLFC mechanism using encoding techniques such as normalization that adjust the values in the range between 0 and 1. The DTLFC mechanism generates the desired features and improves the accuracy rate by 3–9% when compared with the TCI method and by 12–18% when compared with the RMTFS method.

Figure 5 illustrates the processing time based on the attributes. The attribute count ranges from 2, 4, 6, and 8 up to 14. The attribute that has the highest normalized information gain is selected to make the decision in the DTLFC mechanism, so that the processing time is reduced. The decision tree merges the extracted feature set for the medical image point's pixel in the DTLFC mechanism to reduce the processing time by 4–8%

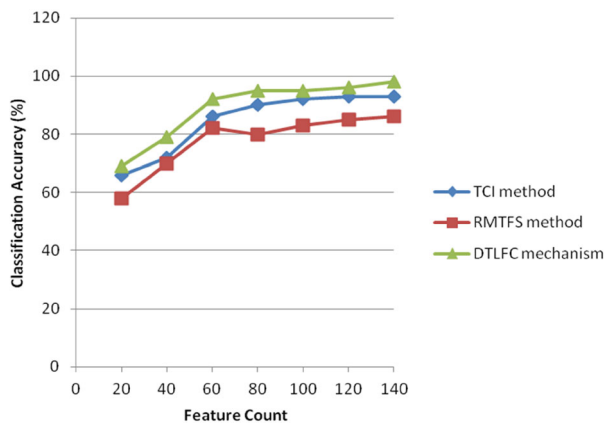


Figure 4. Classification of accuracy measure.

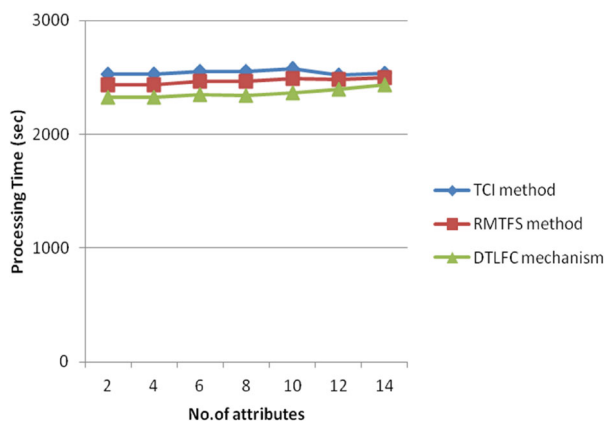


Figure 5. Performance of processing time.

when compared with the TCI method and by 2–5% when compared with the RMTFS method.

Figure 6 depicts the computational cost, measured in terms of milliseconds (ms). The cost varies from 10% to 13% in the TCI method, over the mining operation for useful information from a set of unprocessed and label images. The DTLFC mechanism transforms them into a database table and allows feature extraction algorithms to work on the table, and information is fetched effectively with lower cost. The DTLFC mechanism consumes 20–25% lesser cost when compared with the RMTFS method.

Figure 7 describes the scalability in the DTLFC mechanism, the TCI method and the RMTFS method. The DTLFC mechanism, which uses the concept of information entropy, builds decision trees out of a set of images. When compared with the RMTFS method, the information entropy improves the scalability percentage by a factor of 5. Equation (1) classifies the heart images from the UCI repository effectively and finds the pixels in the x and y directions by smoothening the heart image data points.

Figure 8 depicts the intensity rate based on the attribute count and demonstrates the intensity rate of the dominant attributes based on the tuple count. The tuple count varies from 10, 20, and 30 up to 70. The intensity rate measures the power of the DTLFC mechanism and existing work in terms of watt per metre squared (watt/m^2). The DTLFC mechanism rules are

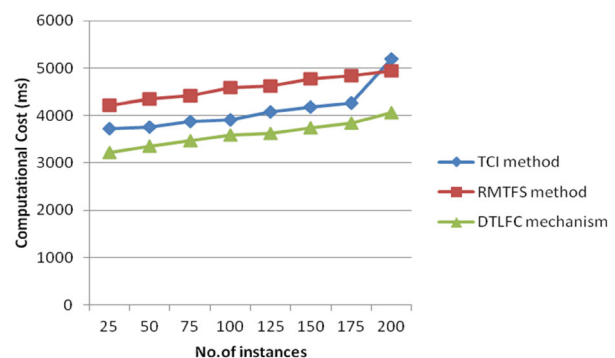


Figure 6. Measure of computational cost.

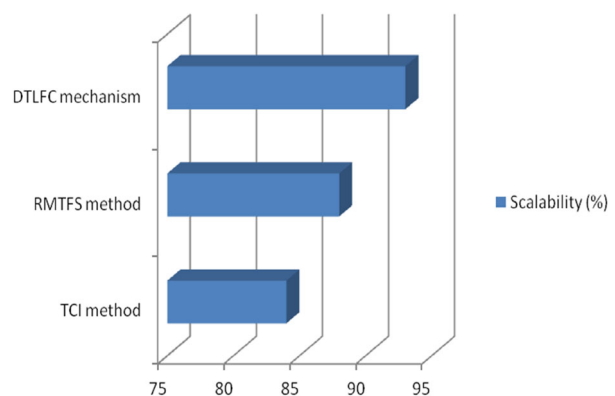


Figure 7. Measure of scalability on various techniques.

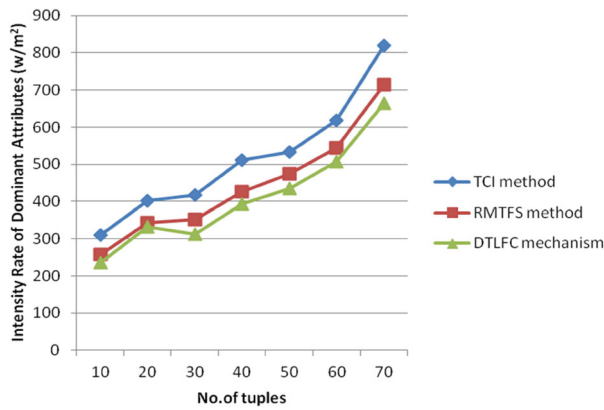


Figure 8. Intensity rate of dominant attributes measure.

generated to interpret and recognize large databases as the tree size is not dependent on the database size. The tuples in the database table pass through the decision tree. The DTLFC mechanism performs the classification operation on the height of the tree, a fixed one to reduce the intensity rate gradually from the TCI and RMTFS methods. The DLFC mechanism has the reduction 18–25% intensity rate attributes in TCI method and 6–11% reduced when compared with the RMTFS method. The dominant attribute of the DTLFC mechanism improves the relationship within the attributes with a lesser power rate.

Finally, the DTLFC mechanism tries to tailor the system with an effective accuracy rate in particular cases. Based on diverse confined pixel information, appropriate image data points are derived easily via the decision rules invoked from the model built. Overall, each of these figures shows results with good accuracy using the decision tree classifier ensemble algorithm, even for relatively SPECT Heart data set images from the UCI repository.

6. Conclusion

The DTLFC performs pixel-wise image feature points extraction and transforms these images into a database table. In the altered database table, each tuple has a feature descriptor from a set of features in combination with the label. The DTLFC mechanism requires label information of image pixels in advance; however, the label information is undetermined. The actual hidden label properties refine the DTLFC mechanism to an unsupervised one for further use. On the other hand, the specialization of the DTLFC mechanism involves the generation of raw image features, the integration of various masks, and the transformation of label image properties. The experimental results of the DTLFC mechanism using the SPECT Heart data set estimate the performance. The mechanism attains maximal filtering efficiency, classification accuracy, minimal processing time, computational cost, and intensity rate of dominant attributes with utmost scalability.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] Lee Y-J, Yeh Y-R, Wang Y-CF. Anomaly detection via online oversampling principal component analysis. *IEEE Trans Knowl Data Eng.* **July 2013**;25(7):1460–1470.
- [2] Parker DR, Gustafson SC, Oxley ME, et al. Development of a Bayesian framework for determining uncertainty in receiver operating characteristic curve estimates. *IEEE Trans Knowl Data Eng.* **January 2010**;22(1):31–45.
- [3] Si S, Tao D, Geng B. Bregman divergence-based regularization for transfer subspace learning. *IEEE Trans Knowl Data Eng.* **July 2010**;22(7):929–943.
- [4] Song Q, Ni J, Wang G. A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Trans Knowl Data Eng.* **January 2013**;25(1):143–156.
- [5] Jiang J-Y, Liou R-J, Lee S-J. A fuzzy self-constructing feature clustering algorithm for text classification. *IEEE Trans Knowl Data Eng.* **March 2011**;23(3):335–349.
- [6] Ramaswamy S, Rose K. Adaptive cluster distance bounding for high-dimensional Indexing. *IEEE Trans Knowl Data Eng.* **June 2011**;23(6):815–830.
- [7] Vanetti M, Binaghi E, Ferrari E, et al. A system to filter unwanted messages from OSN user walls. *IEEE Trans Knowl Data Eng.* **February 2013**;25(2):285–297.
- [8] Papadopoulos S, Wang L, Yang Y, et al. Authenticated multistep nearest neighbor search. *IEEE Trans Knowl Data Eng.* **May 2011**;23(5):641–654.
- [9] Lu J, Ling TW, Bao Z, et al. Extended XML tree pattern matching: theories and algorithms. *IEEE Trans Knowl Data Eng.* **March 2011**;23(3):402–416.
- [10] Shahbazi A, Miller J. Extended subtree: a new similarity function for tree structured data. *IEEE Trans Knowl Data Eng.* **April 2014**;26(4):864–877.
- [11] Angiulli F, Basta S, Lodi S, et al. Distributed strategies for mining outliers in large data sets. *IEEE Trans Knowl Data Eng.* **July 2013**;25(7):1520–1532.
- [12] Yildirim Y, Yazici A, Yilmaz T. Automatic semantic content extraction in videos using a fuzzy ontology and rule-based model. *IEEE Trans Knowl Data Eng.* **January 2013**;25(1):47–61.
- [13] Papadimitriou P, Tsaparas P, Fuxman A, et al. TACI: taxonomy-aware catalog integration. *IEEE Trans Knowl Data Eng.* **July 2013**;25(7):1643–1655.
- [14] Xiang S, Nie F, Zhang C. Semi-supervised classification via local spline regression. *IEEE Trans Pattern Anal Mach Intell.* **November 2010**;32(11):2039–2053.
- [15] Li W-J, Yeung D-Y. MILD: multiple-instance learning via disambiguation. *IEEE Trans Knowl Data Eng.* **January 2010**;22(1):76–89.
- [16] Abbasi A, France S, Zhang Z, et al. Selecting attributes for sentiment classification using feature relation networks. *IEEE Trans Knowl Data Eng.* **March 2011**;23(3):447–462.
- [17] Lee J-G, Han J, Li X, et al. Mining discriminative patterns for classifying trajectories on road networks. *IEEE Trans Knowl Data Eng.* **May 2011**;23(5):713–726.
- [18] Biggio B, Fumera G, Roli F. Security evaluation of pattern classifiers under attack. *IEEE Trans Knowl Data Eng.* **2014**;26(4):984–996.
- [19] Dara RA, Makrehchi M, Kamel MS. Filter-based data partitioning for training multiple classifier systems. *IEEE Trans Knowl Data Eng.* **April 2010**;22(4):508–522.